Exam 1

First Name	Last Name	UID
		U

Honor Code: This exam represents only my own work. I did not give or receive help.

Signature: ______________________________

Partial Credit: There will be partial credit for solution attempts even if not all the mathematical manipulations are completed correctly. It is advisable to attempt every problem.

- You have exactly **2 hours** to complete this exam.
- **No devices are allowed**, including no phones and no calculators. No form of collaboration is allowed.
- You can use the provided formula sheet handout – no other materials are allowed.
- All work to be graded must be included in this document. Submit no extra sheets.
- Answers can be left in terms of sums without having to add or multiply all the terms, unless otherwise specified in the problem. The formula sheet now includes an $\binom{n}{k}$ table for $n = 0, 1, \ldots, 8$ so you are expected to plug in these values when needed.
- There are 8 problems in total, with different point values. Don't get bogged down with any one true/false question. The last problem is **worth fewer points and is conceptually harder:** attempt all other problems before this one.

*** Good Luck! ***

Problem	Points earned	out of	Problem	Points earned	out of
Problem 1		8	Problem 5		16
Problem 2		8	Problem 6		16
Problem 3		16	Problem 7		16
Problem 4		12	Problem 8		8
			Total		100

Problem 1 *8 points*

For each of the following parts, indicate whether the statement is always true or it can be false by clearly writing "True" or "False." Full credit will be given for selecting the correct logical value (even with no explanation). You are welcome to briefly explain your reasoning for partial credit (in case your choice is wrong). Diagrams are welcome. Throughout the problem, you may assume that A, B, and C are events with $\mathbb{P}[A] > 0$, $\mathbb{P}[B] > 0$, and $\mathbb{P}[C] > 0$. The sample space is denoted by Ω.

(a) If $B_1, \ldots, B_n$ are mutually exclusive, then $\mathbb{P}[A] \leq \sum_{i=1}^{n} \mathbb{P}[A|B_i]\,\mathbb{P}[B_i]$

(b) $(A \cap B)^{\mathsf{c}}$, $(A^{\mathsf{c}} \cap B)^{\mathsf{c}}$, $(A \cap B^{\mathsf{c}})^{\mathsf{c}}$, and $(A^{\mathsf{c}} \cap B^{\mathsf{c}})^{\mathsf{c}}$ forms a partition.

(c) Suppose $\mathbb{P}[A|B] > \mathbb{P}[A|B^{\mathsf{c}}]$ with $\mathbb{P}[B] = 1/2$.
Then, given A occurs, we have that B is more likely to occur than B^{c}.

(d) If A and B are conditionally independent given C, then $\mathbb{P}[A \cap B|C] = \mathbb{P}[A \cap B|C^{\mathsf{c}}]$.

Problem 2 *8 points*

For each of the following parts, indicate whether the statement is always true or it can be false by clearly writing "True" or "False." Full credit will be given for selecting the correct logical value (even with no explanation). You are welcome to briefly explain your reasoning for partial credit (in case your choice is wrong). Diagrams are welcome. Throughout the problem, you may assume that X is a discrete random variable with PMF $P_X(x)$ and CDF $F_X(x)$.

(a) $\mathbb{E}[X^2] = \big(\mathbb{E}\big[|X|\big]\big)^2$. (Recall that $|X|$ is the absolute value of X.)

(b) Let X be Binomial(n,p) with $p > 1/2$. Let $B = \{0,1,2,\ldots,m\}$ with $m < n$. Then, the conditional PMF $P_{X|B}(x)$ corresponds to Binomial(m,p). (Note that $\{X \in B\}$ corresponds to the event that you win at most m games out of n.)

(c) Let $A = \big\{x \in R_X : x < \mathbb{E}[X]\big\}$. Then, $\mathbb{E}[X|A] < \mathbb{E}[X]$.

(d) If $\mathbb{E}[(X-1)^2] = \mathbb{E}[X^2] + 1$, then $\mathbb{E}[X] = 0$.

Problem 3 *16 points*

Complete the following quick calculations. **For full credit, you must work out a numerical answer for each requested quantity in this problem.**

(a) Let X be Binomial$(3, \frac{1}{3})$. Compute $\mathbb{P}[X > 1]$ and $\mathbb{P}[2^X < 4]$.

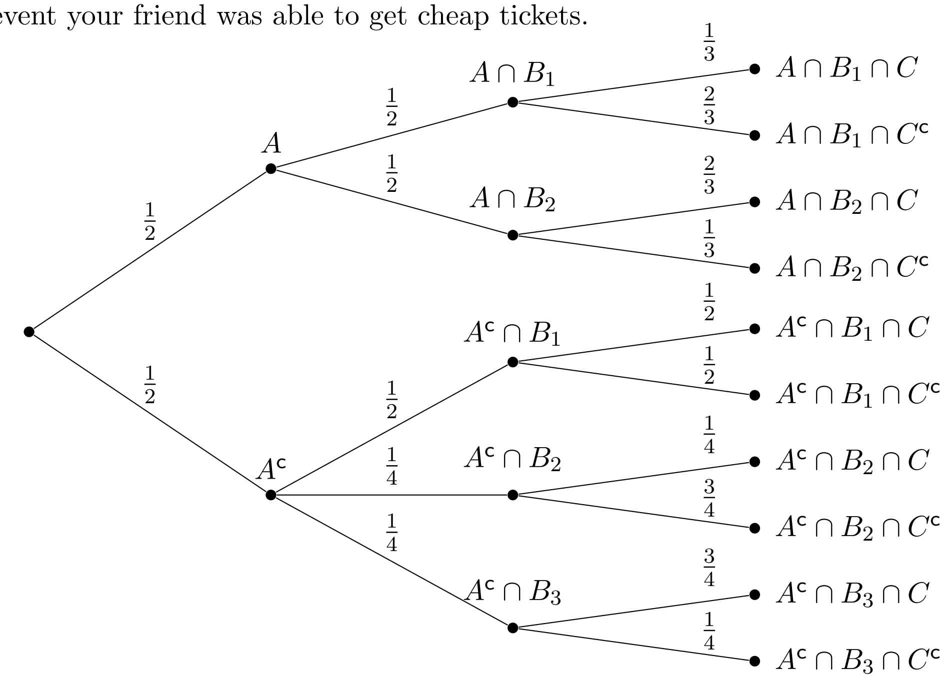
(b) Let X be Geometric$(\frac{1}{2})$. Calculate $\mathsf{Var}[2X + 1]$ and $\mathbb{E}[X^2 - 4]$.

(c) Let X be a Poisson(1) and $A = \{1, 3\}$. Calculate $\mathbb{P}_{X|A}(x)$ and $\mathbb{E}[10^X | A]$.

(d) Let A, B, C be events with $\mathbb{P}[A|C] = \mathbb{P}[B|C] = \mathbb{P}[B|C^{\mathsf{c}}] = \frac{1}{4}$ and $\mathbb{P}[C] = \mathbb{P}[A|C^{\mathsf{c}}] = \frac{1}{2}$. Additionally, let A and B be independent. Calculate $\mathbb{P}[A \cap B]$ and $\mathbb{P}[A \cup B]$.

Problem 4 *12 points*

Your friend attended a concert. Let A be the event it was held in the usual auditorium. Let B_1 be the event that it was your friend's favorite band, B_2 their second favorite, and B_3 their third favorite.
Let C be the event your friend was able to get cheap tickets.



(a) Are the events A and B_1 independent? Justify your answer with calculations.

(b) Given it was their second favorite, what is the probability it was the usual auditorium **and** cheap tickets?

(c) Given that your friend got cheap tickets, what is the probability it was in the usual auditorium?

Problem 5 *16 points*

A pack of 6 widgets contains 3 faulty widgets. You take out 4 widgets. Your formula sheet includes an $\binom{n}{k}$ table so we are expecting exact answers throughout this problem.

(a) How many ways are there to select 4 widgets with exactly 2 faulty?

(b) What is the probability that exactly 2 of the widgets you select are faulty?

(c) Let X be a random variable corresponding to the exact number of faulty widgets you select. Determine the PMF $P_X(x)$ and write it as a case-by-case formula.

(d) What is the average number of faulty widgets you select?

Problem 6 *16 points*

The number of patients checking into a hospital per hour X is well modeled as Poisson(3).
We say that an hour is calm if **at most one** patient checks in.

(a) What is the probability that an hour is calm?

(b) Let B be the event that an hour is calm. Determine the conditional PMF $P_{X|B}(x)$ and write it down as a case-by-case formula.

(c) Consider now the following scenario. Out of 4 calm hours, we count the total number of patients Y that check in. What kind of random variable is Y? Don't forget the parameters.

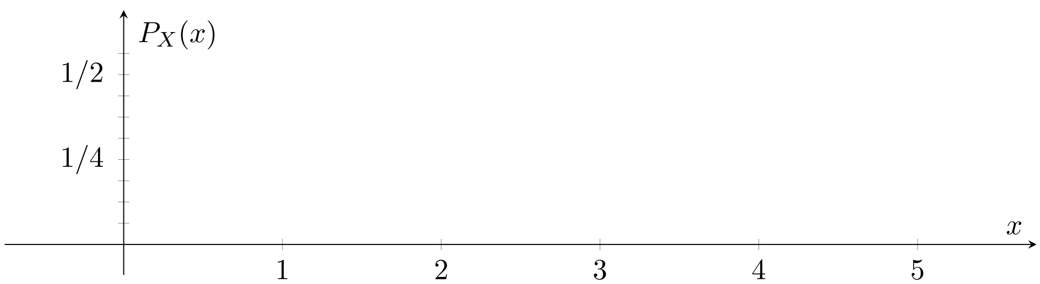
(d) What is the probability that at least one patient checks in across 4 calm hours?

Problem 7 *16 points*

A cereal company includes prizes in some cereal boxes with probability $\frac{1}{2}$, independently across boxes. Let X be the number of boxes purchased until the first prize. Your friend is working as an intern with this company, and they designed a survey with the following questions and preliminary collected data.

How many boxes until first prize?	# Responses	(a) Histogram Values	(d) Predicted Probabilities
1	8		
2	3		
3 or more	5		

(a) Using the normalized counts (like we did in the labs), determine the histogram values for this dataset. Fill in your values in both the table above under column (a) Histogram Values and as a bar plot in the figure below. **Your friend has decided you should plot "3 or more" at $x = 3$.**



(b) What kind of random variable is X? Don't forget the parameters. Draw its PMF as a stem plot on the figure above for all x values on the x-axis.

(c) Your friend believes that, if they collect more data with the existing survey questions, the histogram and PMF will eventually agree. Do you agree, yes or no? Support your answer with probability concepts.

(d) Using the PMF for the random variable X, determine the predicted values for the survey questions responses, assuming your friend collects much more data. Write your answers in column (d) Predicted Probabilities in the table above.

Problem 8 *8 points*

Your company has created a collection of N machine learning algorithms for classifying images as either "cat" or "dog." Your dataset contains more dog images than cat images, and you may assume the probability of getting a dog image is $\frac{2}{3}$ and, otherwise, you get a cat image. Each algorithm classifies a cat image correctly as "cat" with probability p and a dog image correctly as "dog" with probability q. The decisions of each algorithm can be modeled as independent of one another. Your insight is that, to improve the overall performance, you can run all N algorithms on an image, and use the majority vote as the output. For instance, if $N = 3$ and the algorithms output the guesses "cat," "dog," "cat," you will guess this is a cat image. You may assume throughout the problem that N is odd.

(a) What is the probability your majority-voting algorithm is correct? (It is OK to leave your answer in terms of summations.)

(b) Given that your majority-voting algorithm was correct, what is the probability it was actually a cat image for $N = 3$, $p = 2/3$, $q = 1/2$? (For full credit, please provide a numerical answer with all terms simplified.)